Supplementary Appendix

This appendix has been provided by the authors to give readers additional information about their work.

Supplement to: Köser CU, Holden MTG, Ellington MJ, et al. Rapid whole-genome sequencing for investigation of a neonatal MRSA outbreak. N Engl J Med 2012;366:2267-75.

Supplement for "A neonatal MRSA outbreak investigation using rapid whole genome sequencing"

Claudio U. Köser*, Matthew T.G. Holden*, Matthew J. Ellington*, Edward J.P. Cartwright*, Nicholas M. Brown, Amanda L. Ogilvy-Stuart, Li Yang Hsu, Claire Chewapreecha, Nicholas J. Croucher, Simon R. Harris, Mandy Sanders, Mark C. Enright, Gordon Dougan, Stephen D. Bentley, Julian Parkhill, Louise J. Fraser, Jason R. Betley, Ole B. Schulz-Trieglaff, Geoffrey P. Smith, Sharon J. Peacock
*contributed equally

Table of contents

Supplementary Methods	3
Figure S1	7
Table S1	8
Table S2	10
Table S3	11
Table S4	12
References	13

Supplementary Methods

Bacterial isolation, identification and drug susceptibility testing in the routine diagnostic laboratory

MRSA in blood cultures were detected using an automated blood culture system (BacT/ALERT 3D, bioMérieux, Marcy l'Etoile, France) followed by sub-culture on Columbia agar with horse blood (Oxoid, Basingstoke, UK) for 18-24 hours at 37°C. MRSA screening swabs were directly plated on Brilliance MRSA 2 agar plates (Oxoid). *S. aureus* identification was achieved using the Pastorex Staph-Plus (BioRad, Marnes la Coquette, France) and coagulase tests (E&O Laboratories, Bonnybridge, UK and Oxoid). Susceptibility testing to eight antimicrobial drugs reported to clinicians (cefoxitin (CXT), erythromycin (ERY), ciprofloxacin (CIP), gentamicin (GEN), tetracycline (TET), rifampicin (RIF), fusidic acid (FUS), and mupirocin (MUP)) was performed using disk susceptibility testing according to BSAC criteria on ISO sensitest agar (Oxoid), with the exception of isolate 19B which was a small colony variant and tested on ISO blood + NAD plates (Oxoid).

Additional growth conditions and phenotypic testing

Following isolation from clinical samples, all MRSA isolates were stored at -80°C. A secondary frozen stock was made from this and used for all experimental work described in this study. The primary stock was streaked onto brilliance MRSA 2 agar plates (Oxoid) with the exception of isolate 19B, which was an auxotrophic small colony variant that grew poorly on this medium and was therefore grown on Columbia agar with horse blood (Oxoid). A single colony was inoculated into a brain heart infusion broth (BHI) (Colindale, UK), incubated overnight at 37°C in air, and secondary frozen stocks prepared. Extended antimicrobial susceptibility testing to five additional antibiotics (clindamycin (CLIN), kanamycin (KAN), tobramycin (TOB), trimethroprim (TMP), and linezolid (LIN)) was determined for the 14 MRSA isolates that were sequenced as part of this study using the *Staphylococcus* spp. kit with the Vitek 2 system version 4.02 (bioMérieux) according to EUCAST 2008 guidelines. This also confirmed the initial eight susceptibility test results, above.

Isolate 6C had a higher than expected number of SNPs compared with other isolates cultured during a putative outbreak. The finding that this isolate had a genetic mutation in *mutS* consistent with being a hypermutator (see main text) led us to confirm the presence of a hypermutator phenotype. Mutation frequencies for 6C and 7C, its closest relative, were determined and compared with *S. aureus* RN4220 as negative and RN4220Δ*mutS* as

positive controls as previously described, with the exception that 6C was allowed to grow for 50% longer to compensate for its slow growth rate.[1]

DNA isolation

A total of 14 MRSA isolates (seven associated with the NICU outbreak isolate and seven associated with bacteremias in patients on other wards in the same hospital) were chosen for whole genome sequencing. Cells from a 9 ml BHI overnight culture (37°C) were digested in 20mM EDTA (Sigma-Aldrich, Steinheim, Germany), 0.2 mg/ml lysostaphin (Sigma-Aldrich), and 40.0 mg/ml lysozyme (Sigma-Aldrich) for 1 hour at 37°C. DNA was then extracted using a QIAamp DNA Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions.

DNA sequencing

The ST22 reference isolate used in this study was HO 5096 0412, a representative of EMRSA-15 that was isolated from a neonate with fatal infection in Suffolk, UK in March 2005. This genome had been sequenced prior to this study, and was utilized to map the genetic diversity of ST22 isolates from Addenbrooke's Hospital. In brief, this genome was completely sequenced, assembled, finished and annotated, as described previously.[2,3] The genome was sequenced to approximately 11-fold coverage from pMAQ1Sac_BstXI (insert size 2-12 kb) genomic shotgun libraries using big-dye terminator chemistry on ABI3730 automated sequencers. End sequences from large insert fosmid libraries in pCC1Fos (insert size 38-42 kb) were used as a scaffold. All repeat regions were bridged by read-pairs or end-sequenced polymerase chain reaction (PCR) products. The HO 5096 0412 genome consists of a single circular chromosome of 2,832,299 bp (deposited under the accession number HE681097 at the European Nucleotide Archive (ENA, http://www.ebi.ac.uk/ena/)) and a plasmid of 2,473 bp (identical to CP002148).

For the MRSA isolates from Addenbrooke's Hospital, DNA was quantified using the Quant-iT dsDNA High-Sensitivity Assay Kit with the Qubit Fluorometer (Invitrogen), and 50 ng of each sample was prepared for sequencing using a Nextera DNA Sample Prep Kit (Epicentre) with indexing. As recommended in the manufacturer's protocol, to reduce the number of fragments smaller than 300 bp, a clean-up step with Agencourt AMPure XP beads was carried out instead of the Zymo DNA Clean and Concentrator-5, with the minor alteration of 0.6x beads instead of 0.7x beads. The resulting libraries were quantified using the Qubit High sensitivity kit, and the library size assessed using an Agilent 2100 bioanalyzer with the High Sensitivity DNA Kit. Each of 12 indexed libraries were diluted to 2 nM in buffer EB

(Qiagen), and then pooled in equimolar ratio (the maximum number for multiplexed sequencing was 12) and the remaining two libraries diluted and pooled together as before. The total time for the library preparation was approximately 3 hrs including initial quantification (20min), library generation (1.5hr), and final library QC (1hr). The pooled libraries were denatured using sodium hydroxide and diluted to 6-8 pM in buffer HT1 as described in the standard Illumina protocol, and sequenced in two Illumina MiSeq runs (1st run: 12 isolates, 2nd run: 2 isolates) following standard procedures. In brief, the sample was loaded in the sample well of a MiSeq consumable cartridge and loaded onto the MiSeq instrument. Runs were initiated for 2 x 150 bases of "sequencing-by-synthesis" sequencing, including on-board clustering, paired-end preparation, and sequencing of the indexes. On completion of the runs, data was automatically demultiplexed on the instrument to generate fastq files for each isolate that were used for subsequent isolate comparison. The genome data has been deposited at ENA (accession number ERP001256).

Sequence analysis

The multilocus sequence types (MLST) of the 14 isolates were assigned from the sequence data as described by Croucher et al.[4] Reads were mapped to the chromosome of *S. aureus* HO 5096 0412 as paired end reads with an insert size between 50 and 400 bp using SMALT (http://www.sanger.ac.uk/resources/software/smalt/). Single nucleotide polymorphisms (SNPs) were identified as described in Harris et al.,[5] and indel regions as described by Croucher et al.[4] (Table S3). The phylogeny was constructed on the basis of core SNPs using RAxML as described in Harris et al.[5] The HO 5096 0412 core genome was defined by comparative genomic comparisons with other complete *S. aureus* genomes in the public databases (for details of the excluded accessory regions, see Table S4). To increase the resolution in distinguishing the ST1 isolates, 16B and 17B, the sequence data for both isolates were mapped against the chromosome of the ST1 isolate MSSA476 (accession number BX571857).[3]

Resistome and toxome

The accession numbers of the genes tested for in the resistome and toxome are shown in Table S2. Reads were mapped to the resistome and toxome pseudomolecules as paired end reads with an insert size between 50 and 400 bp using BWA.[6] In Figure 2B, red indicates >10 reads, blue indicates <10 reads with a window size of 10 bp. Antimicrobial resistance conferred by SNPs in components of the core chromosome were identify by manual curation; substitutions were compared to antimicrobial resistance mutations in *S. aureus* proteins detailed in the literature. For all the isolates, LIN susceptibility was checked

by searching for the presence of the *cfr* gene (accession number CFR_STAAU[7]), and substitutions in core genes associated with LIN resistance.[8]

Figure S1 Rooted parsimony tree of ST22 isolates that displays SNPs on each branch.

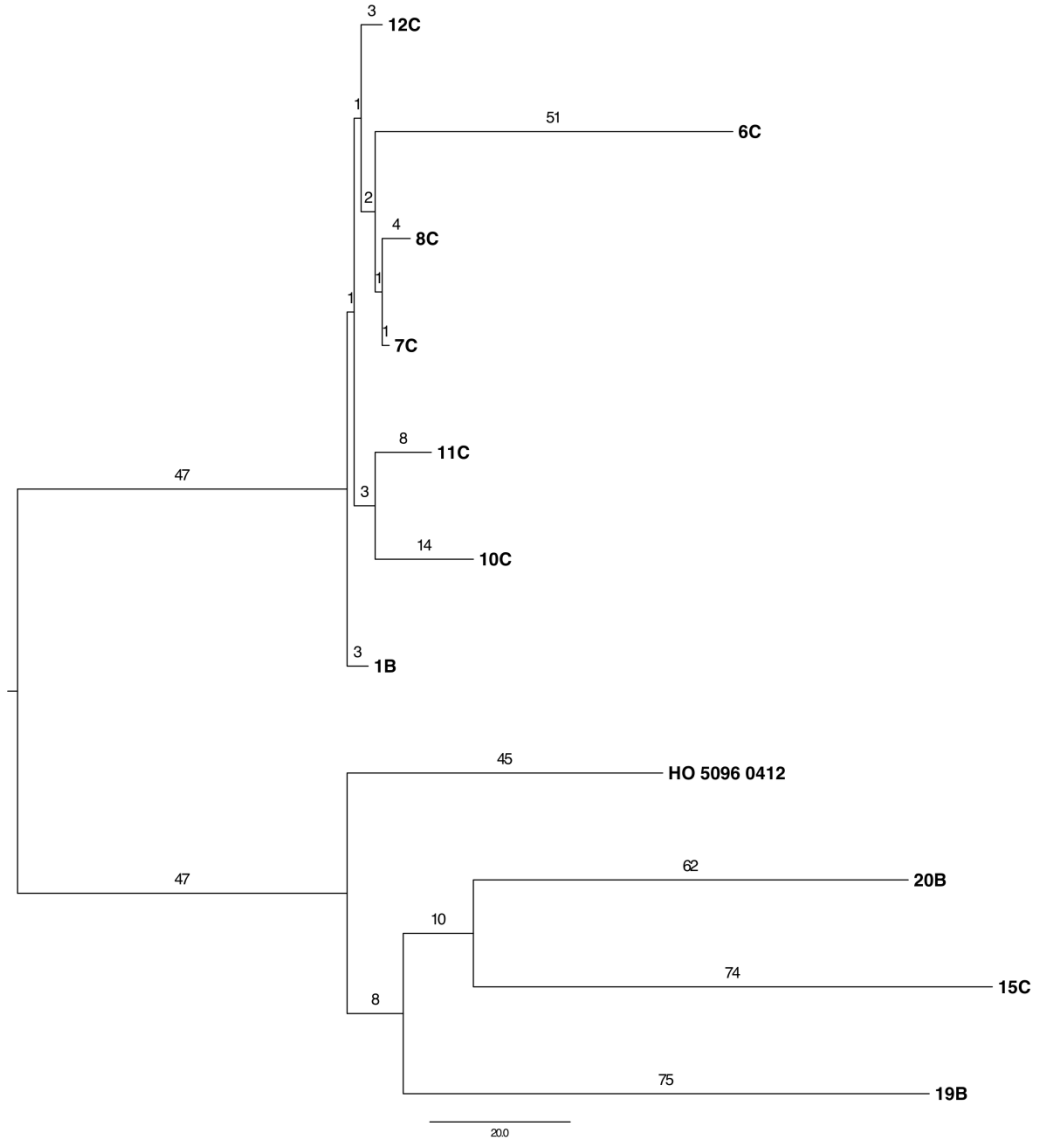

Table S1 Details of 51 SNPs that were unique to isolate 6C.

Position in reference	Base substitution	CDS	CDS Systematic ID in reference
Intergenic SNPs			
411648	C->T		
531396	T->C		
551418	C->T		
752914	A->G		
2178094	C->T		
2376535	C->G		
2418772	G->A		
2453996	C->T		
Synonymous SNPs			
66220	G->A		
229444	C->T		
1333266	G->A		
1358124	G->C		
2145698	G->T		
Non-synonymous SNPs			
57247	G->A	conserved hypothetical protein	SAEMRSA1500430
68396	G->C	TetR family regulatory protein	SAEMRSA1500520
100293	C->T	putative regulatory protein	SAEMRSA1500780
254859	C->A	putative PTS transport system, IIBC component	SAEMRSA1502010
263260	C->T	putative zinc-binding dehydrogenase	SAEMRSA1502090
280525	G->T	putative nitric oxide reductase	SAEMRSA1502230
343442	G->A	NADH:flavin oxidoreductase / NADH oxidase family protein	SAEMRSA1502780
538237	G->A	putative stress response-related Clp ATPase	SAEMRSA1504510
588244	C->A	bone sialoprotein-binding protein	SAEMRSA1504900
626242	G->C	conserved hypothetical protein	SAEMRSA1505290
806811	G->A	glyceraldehyde 3-phosphate dehydrogenase 1	SAEMRSA1506990
877349	C->T	hypothetical protein	SAEMRSA1507750
906871	C->A	putative membrane protein	SAEMRSA1507990
1044514	C->G	dihydrolipoamide dehydrogenase	SAEMRSA1509260
1095952	C->A	putative excinuclease ABC subunit C	SAEMRSA1509760
1177464	T->A	conserved hypothetical protein	SAEMRSA1510590
1222127	G->A	prolyl-tRNA synthetase	SAEMRSA1510960
1387158	A->T	phosphate-binding lipoprotein	SAEMRSA15125

			30
1418036	G->C	putative membrane protein	SAEMRSA1512800
1688784	G->A	conserved hypothetical protein	SAEMRSA1515370
1747359	C->G	threonyl-tRNA synthetase	SAEMRSA1515930
1794131	G->A	conserved hypothetical protein	SAEMRSA1516290
1807212	A->T	haptoglobin-binding surface anchored protein*	SAEMRSA1516390
1807213	G->T	haptoglobin-binding surface anchored protein*	SAEMRSA1516390
1824217	C->T	FtsK/SpoIIIE family protein	SAEMRSA1516490
1866196	C->G	conserved hypothetical protein (fragment)	SAEMRSA1516830
1877043	G->T	aldo/keto reductase family protein	SAEMRSA1516960
2012913	C->A	putative prephenate dehydratase	SAEMRSA1518250
2019150	G->A	putative aldehyde dehydrogenase	SAEMRSA1518300
2109350	G->A	autoinducer sensor protein response regulator protein	SAEMRSA1519470
2471809	C->G	hypothetical protein	SAEMRSA1522920
2498986	G->A	IgG-binding protein	SAEMRSA1523170
2535632	C->T	putative glycine betaine/carnitine/choline transport ATP-binding protein	SAEMRSA1523480
SNOP SNPs			
1116321	C->T	putative carbamate kinase	SAEMRSA1510010
1262441	G->A	DNA mismatch repair protein MutS	SAEMRSA1511300
1469705	G->A	hypothetical protein	SAEMRSA1513030
2308571	T->A	30S ribosomal protein S9	SAEMRSA1521160
2382938	G->T	putative bifunctional protein	SAEMRSA1522070

*SNPs that affect the same amino acid.

Table S2 Sequences used in the resistome and toxome pseudomolecules that were mapped with the MiSeq data.

Gene	Accession	Reference
mecA	X52592	[9]
ermA	P06699	[10]
ermC	P13978	[11]
aacA-aphD	P14507	[12]
aadD	P05057	[13]
ant1	P0A0D1	[14]
tetK	P02983	[15]
dfrG	C7C2U7	[16]
fusC	Q6GD50	[17]
ileS-2	P41972	[18]
sea	M18970	[19]
seb	M11118	[19]
sec	X05815	[19]
sed	M28521	[19]
see	M21319	[19]
seg	AF064773	[19]
seh	U11702	[19]
sei	AF064774	[19]
sej	AF053140	[19]
eta	P09331	[20]
etb	AAA26628	[20]
etd	BAC22944	[20]
tst	J02615	[19]
lukS-PV	X72700	[21]
lukF-PV	X72700	[21]

Table S3 Summary results of mapping MiSeq data to the ST22 reference (HO 5096 0412).

Isolate ID	ST	Bases mapped	Mapping coverage (times)	Number of SNPs[1]
1B	22	2,744,482	33	215
6C	22	2,755,142	40	271
7C	22	2,754,560	40	218
8C	22	2,756,942	39	219
10C	22	2,758,798	42	232
11C	22	2,742,777	33	217
12C	22	2,750,508	37	216
14C	5	2,561,203	39	39,913
15C	22	2,763,374	46	175
16B	1	2,504,465	23	38,706
17B[2]	1	2,561,309	56	39,985
18B	36	2,607,209	37	50,620
19B[2]	22	2,769,080	111	376
20B	22	2,752,049	34	130

[1]Including SNPs in the accessory genome.

[2]Sequenced in the second run.

Table S4 Accessory regions of the *S. aureus* HO 5096 0412 reference chromosome, excluded in the designation of core SNPs.

Start	End	Feature
34163	51525	SCC*mec* type IV
80765	83026	IS element
137322	138041	IS element
141474	142274	IS element
315945	318143	IS element
818385	820574	IS element
935317	936294	IS element
938088	938798	IS element
977393	979583	IS element
1104852	1107038	IS element
1252874	1254311	IS element
1283817	1286016	IS element
1328804	1330543	IS element
1361018	1375548	ICE6013
1520143	1566314	Prophage φSa2 (HO 5096 0412)
1567574	1569770	IS element
1852864	1853851	IS element
1861820	1863920	Arsenic resistance operon of plasmid origin
1874756	1876610	IS element
1890687	1892984	IS element
1943176	1944547	IS element
1944548	1945123	IS element
2021753	2029411	Tn*552*-like transposon
2042948	2087965	Prophage φSa3 (HO 5096 0412)
2136387	2138610	IS element
2509256	2510560	IS element
2568107	2570285	IS element
2700680	2701410	IS element
2825393	2826315	IS element

References

1. O'Neill AJ, Chopra I. Insertional inactivation of *mutS* in *Staphylococcus aureus* reveals potential for elevated mutation frequencies, although the prevalence of mutators in clinical isolates is low. J Antimicrob Chemother 2002;50:161-9.
2. Parkhill J, Achtman M, James KD, et al. Complete DNA sequence of a serogroup A strain of *Neisseria meningitidis* Z2491. Nature 2000;404:502-6.
3. Holden MT, Feil EJ, Lindsay JA, et al. Complete genomes of two clinical *Staphylococcus aureus* strains: evidence for the rapid evolution of virulence and drug resistance. Proc Natl Acad Sci U S A 2004;101:9786-91.
4. Croucher NJ, Harris SR, Fraser C, et al. Rapid pneumococcal evolution in response to clinical interventions. Science 2011;331:430-4.
5. Harris SR, Feil EJ, Holden MT, et al. Evolution of MRSA during hospital transmission and intercontinental spread. Science 2010;327:469-74.
6. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 2009;25:1754-60.
7. Toh SM, Xiong L, Arias CA, et al. Acquisition of a natural resistance gene renders a clinical strain of methicillin-resistant *Staphylococcus aureus* resistant to the synthetic antibiotic linezolid. Mol Microbiol 2007;64:1506-14.
8. Locke JB, Hilgers M, Shaw KJ. Novel ribosomal mutations in *Staphylococcus aureus* strains identified through selection with the oxazolidinones linezolid and torezolid (TR-700). Antimicrob Agents Chemother 2009;53:5265-74.
9. Ryffel C, Tesch W, Birch-Machin I, et al. Sequence comparison of *mecA* genes isolated from methicillin-resistant *Staphylococcus aureus* and *Staphylococcus epidermidis*. Gene 1990;94:137-8.
10. Murphy E. Nucleotide sequence of *ermA*, a macrolide-lincosamide-streptogramin B determinant in *Staphylococcus aureus*. J Bacteriol 1985;162:633-40.
11. Catchpole I, Thomas C, Davies A, Dyke KG. The nucleotide sequence of *Staphylococcus aureus* plasmid pT48 conferring inducible macrolide-lincosamide-streptogramin B resistance and comparison with similar plasmids expressing constitutive resistance. J Gen Microbiol 1988;134:697-709.
12. Rouch DA, Byrne ME, Kong YC, Skurray RA. The *aacA-aphD* gentamicin and kanamycin resistance determinant of Tn*4001* from *Staphylococcus aureus*: expression and nucleotide sequence analysis. J Gen Microbiol 1987;133:3039-52.
13. Matsumura M, Katakura Y, Imanaka T, Aiba S. Enzymatic and nucleotide sequence studies of a kanamycin-inactivating enzyme encoded by a plasmid from thermophilic bacilli in comparison with that encoded by plasmid pUB110. J Bacteriol 1984;160:413-20.
14. Murphy E. Nucleotide sequence of a spectinomycin adenyltransferase AAD(9) determinant from *Staphylococcus aureus* and its relationship to AAD(3") (9). Mol Gen Genet 1985;200:33-9.
15. Mojumdar M, Khan SA. Characterization of the tetracycline resistance gene of plasmid pT181 of *Staphylococcus aureus*. J Bacteriol 1988;170:5522-8.
16. Kadlec K, Schwarz S. Identification of a novel trimethoprim resistance gene, *dfrK*, in a methicillin-resistant *Staphylococcus aureus* ST398 strain and its physical linkage to the tetracycline resistance gene *tet*(L). Antimicrob Agents Chemother 2009;53:776-8.
17. O'Neill AJ, McLaws F, Kahlmeter G, Henriksen AS, Chopra I. Genetic basis of resistance to fusidic acid in staphylococci. Antimicrob Agents Chemother 2007;51:1737-40.
18. Chalker AF, Ward JM, Fosberry AP, Hodgson JE. Analysis and toxic overexpression in *Escherichia coli* of a staphylococcal gene encoding isoleucyl-tRNA synthetase. Gene 1994;141:103-8.
19. Monday SR, Bohach GA. Use of multiplex PCR to detect classical and newly described pyrogenic toxin genes in staphylococcal isolates. J Clin Microbiol 1999;37:3411-4.
20. Yamaguchi T, Nishifuji K, Sasaki M, et al. Identification of the *Staphylococcus aureus etd* pathogenicity island which encodes a novel exfoliative toxin, ETD, and EDIN-B. Infect Immun 2002;70:5835-45.
21. Lina G, Piémont Y, Godail-Gamot F, et al. Involvement of Panton-Valentine leukocidin-producing *Staphylococcus aureus* in primary skin infections and pneumonia. Clin Infect Dis 1999;29:1128-32.